

融合空间注意力机制的行车障碍预测网络

雷俊锋, 贺睿, 肖进胜

引用本文:

雷俊锋, 贺睿, 肖进胜. 融合空间注意力机制的行车障碍预测网络[J]. *光学精密工程*, 2020, 28(8): 1850–1860.

LEI Jun-feng, HE Rui, XIAO Jin-sheng. Driving obstacles prediction network merged with spatial attention[J]. *Optics and Precision Engineering*, 2020, 28(8): 1850–1860.

在线阅读 View online: <https://doi.org/10.3788/OPE.20202808.1850>

您可能感兴趣的其他文章

Articles you may be interested in

利用卷积神经网络的自动驾驶场景语义分割

Autonomous driving semantic segmentation with convolution neural networks

光学精密工程. 2019, 27(11): 2429–2438 <https://doi.org/10.3788/OPE.20192711.2429>

循环神经网络多标签航空图像分类

Recurrent neural network multi-label aerial images classification

光学精密工程. 2020, 28(6): 1404–1413 <https://doi.org/10.3788/OPE.20202806.1404>

基于语义分割的复杂场景下的秸秆检测

Straw detection algorithm based on semantic segmentation in complex farm scenarios

光学精密工程. 2020, 28(1): 200–211 <https://doi.org/10.3788/OPE.20202801.0200>

基于Inception-SSD算法的零件识别

Target part recognition based Inception-SSD algorithm

光学精密工程. 2020, 28(8): 1799–1809 <https://doi.org/10.3788/OPE.20202808.1799>

非合作纹理目标单目位姿计算

Monocular pose calculation of non-cooperation textured object

光学精密工程. 2020, 28(8): 1775–1784 <https://doi.org/10.3788/OPE.20202808.1775>

文章编号 1004-924X(2020)08-1850-11

融合空间注意力机制的行车障碍预测网络

雷俊锋*, 贺睿, 肖进胜
(武汉大学电子信息学院, 湖北武汉 430072)

摘要:针对现有行车障碍预测方法存在目标单一性、预测速度慢和准确性不佳等问题,提出一种融合空间注意力机制的卷积神经网络 Coll-Net 以及基于 Coll-Net 的车速控制和障碍方向判定策略。模拟驾驶员通过视觉信息判断障碍的机制,以单目视觉图像作为输入,首先对图像做预处理得到感兴趣区域,然后利用残差块网络提取区域内的空间特征;采用空间注意力机制对特征通道上的原始特征进行重新标定,获得通道权重;再将通道权重归一化后加权到通道对应的空间特征上,以此挑选关键特征,最后送入全连接层和 Sigmoid 函数中生成预测概率。行车根据障碍预测概率实时确定行车速度并根据多窗口的概率预测值判定障碍方向。实验表明,Coll-Net 模型的障碍预测准确率达到 96.01%,F1-score 达到 0.915,模型推理时间仅需 24 ms,能够实时检测车辆、行人、护栏、墙体等多种障碍物,并且在低对比度光照环境下仍表现出良好的预测能力,基于 Coll-Net 的车速控制和障碍方向判定策略在 Udacity Self-Driving 数据集上表现出强有效性。

关键词:障碍预测;单目视觉;深度学习;注意力机制;智能驾驶

中图分类号:TP29,U495 **文献标识码:**A **doi:**10.3788/OPE.20202808.1850

Driving obstacles prediction network merged with spatial attention

LEI Jun-feng*, HE Rui, XIAO Jin-sheng

(Electronic Information School, Wuhan University, Wuhan 430072, China)

* Corresponding author, E-mail: jflei@whu.edu.cn

Abstract: To address the limited detection targets, slow processing speed, and low accuracy of existing methods for driving obstacle prediction, this paper proposed an improved convolutional neural network called Coll-Net merged with spatial attention, a suitable speed control policy, and an obstacle direction determination method based on Coll-Net. Coll-Net imitated the vision mechanism of judging obstacles during driving, preprocessed the input monocular vision images to obtain the region of interest, and extracted the spatial features using a deep residual network framework. After collecting the spatial features, Coll-Net recalibrated the original features on the spatial feature channels by using the mechanism of spatial attention, which evaluated the features of every channel, improved the important ones, and then rescaled the weights of every channel and assigned the normalized weights to the corresponding spatial features in order to select critical features. The output feature map is connected by a fully connected layer; then, a normalized obstacle probability range of 0 to 1 is generated by a sigmoid function. Moreover, this paper proposes a driving policy, that controls the driving speed and predicts the obstacle direction according to the generated probability by Coll-Net. Experiment results indicate

收稿日期:2020-04-27;修订日期:2020-05-22.

基金项目:国家自然科学基金资助项目(No. 41771457)

that Coll-Net prediction accuracy on standard datasets reaches 96.01% and the f1 score reaches 0.915. Coll-Net performs well in detecting diverse obstacles such as cars, pedestrians, guardrails, and walls in real time(24 ms for inference), as well as in low-contrast conditions. Moreover, the driving policy based on Coll-Net is validated using Udacity Self-Driving datasets.

Key words: obstacles prediction; monocular vision; deep learning; attention mechanism; intelligent driving

1 引言

先进辅助驾驶系统(Automatic Data Acquisition System, ADAS)是智能交通领域内的研究热点。ADAS 实时收集和处埋车内外环境数据并做出反馈,从而让驾驶者迅速察觉可能发生的危险。前碰撞预警系统(Forward Collision Warning System, FCWS)是 ADAS 系统的一部分,通过车前传感器感知环境并获得障碍信息,在碰撞前发出紧急碰撞警告。早期的机器视觉研究者通常采用目标检测的方法,定位图像中的常见障碍物^[1-2];YOLO(You Only Look Once),RCNN(Regions with Convolutional Neural Networks)等^[3-4]目标检测网络的出现,在车辆和行人目标的检测上有较好的效果^[5],但对于树木、墙壁等不常见障碍物的检测效果不好;由于全卷积网络的出现^[6-7],也有研究者采用语义分割的方法进行自动驾驶场景的识别^[8],但是硬件开销大且预测速度慢,无法达成实时检测的任务。

伴随着从 AlexNet^[9]到 EfficientNet^[10]的目标分类网络的发展,部分研究人员提出将障碍预测从检测车辆、行人等具体目标,转变为直接预测前方障碍的概率。基于该思想,Giusti 等^[11]提出使用神经网络将输入图分为可继续行驶类与不可行驶类;Richter^[12]研究了 DNN 在障碍预测中的场景不确定性问题。这些研究使得障碍预测不再局限于车和行人目标的检测,对大部分障碍物都有一定的预测能力,但预测准确性不佳,不能确定障碍物方位,缺乏对行车控制的研究,仍不能满足行车避障需求。针对以上不足,本文提出一种融合空间注意力机制的卷积神经网络 Coll-Net(意 Collision Network),与现有算法相比,Coll-Net 在满足实时运行的前提下较大提升了障碍预测算

法的准确性,并对护栏、树丛等障碍均具有预测能力;同时,提出了基于 Coll-Net 的行车策略,辅助系统进行障碍方位的判定以及车速控制,以满足智能驾驶中的避障需求。

2 基于驾驶员视觉信息的障碍预测

行车时驾驶员判断前方障碍的重点是障碍的视觉特征而非障碍物与行车的绝对距离,比如针对出现在前方路面的电动车,驾驶员首先根据视觉中电动车较远判断不会发生碰撞,然后根据电动车由远及近的视觉信息判断可能与电动车发生碰撞,如图 1 所示。据此,本文将行车视野图划分为不发生障碍碰撞与可能碰撞两类,并设计 Coll-Net 网络对前方出现障碍的概率进行预测。



图 1 行车视野

Fig. 1 Driving visual field

本文构建 Coll-Net 行车障碍预测系统如图 2,其工作原理是:单目摄像头捕捉前方视野图像,这些图像经预处理^[13]后送入 Coll-Net,经池化下采样处理以及残差网络骨架提取空间特征后,由重标定网络进行通道特征加权以挑选出重要特征并生成空间特征图,这些图最后由全连接层处理并归一化为障碍概率。Coll-Net 的预测值供处理

器做出行车决策,系统将根据障碍概率进行车速控制,并结合图像中多窗口的预测值进行障碍方向判定。

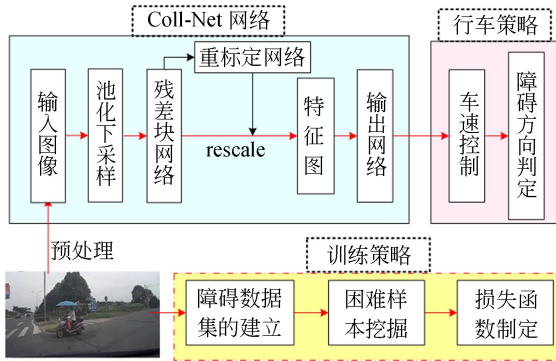


图 2 Coll-Net 行车障碍预测系统构成

Fig. 2 Framework of Coll-Net obstacle prediction system

3 障碍预测网络设计

3.1 Coll-Net 网络设计

3.1.1 整体网络框架

Coll-Net 是由残差网络骨架 Residual Network^[14]根据空间注意力机制改良而得到的卷积神经网络,如图 3 所示,系统工作时,采集图像首先预处理成 200×200 的单通道图像,然后输入到含有 32 个尺寸为 5×5 的卷积核且步长为 2 的卷积层,再采用尺寸为 3×3 且步长为 2 的最大池化层对其进行下采样而生成 32 张尺寸为 50×50 的特征图,然后在同一批次图像中完成特征图的批标准化(BN)和非线性化(PReLU)操作。采样层输出的特征图尺寸为:

$$L_{out} = (L_{in} - W_{kernel}) / S_{stride} + 2, \quad (1)$$

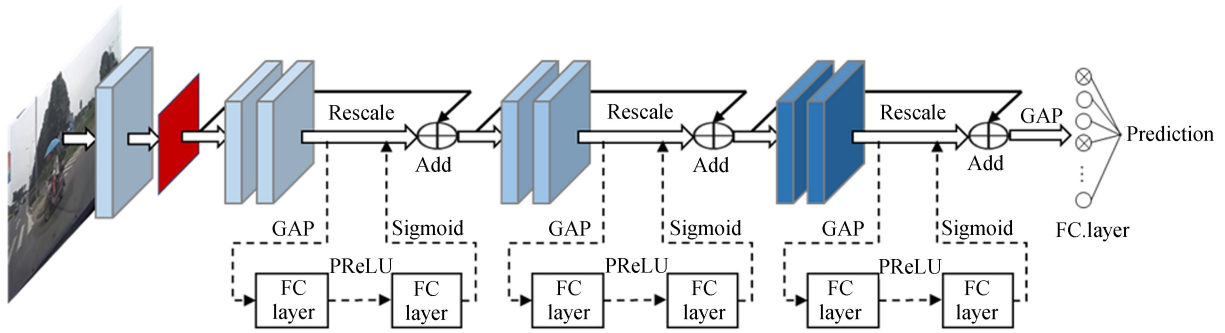


图 3 Coll-Net 深度神经网络结构

Fig. 3 Coll-Net deep neural network structure

其中: L_{in} , L_{out} 表示输入与输出图尺寸, S_{stride} , W_{kernel} 表示步长与核尺寸;然后,特征图依次经过 3 个结构相同但通道数和步长不同的改良残差卷积块,得到空间注意力机制挑选后的关键特征;最后,采用全局平均池化方法(Global Average Pooling, GAP)和全连接层进行处理后,使用 Sigmoid 函数输出归一化的障碍预测概率。表 1 给出了 Coll-Net 行车障碍预测网络结构涉及的操作,表中 Rescale 表示是否使用通道注意力机制进行重标定,Add 表示与上个卷积块的输出进行连接, S 表示步长,FC 表示全连接层处理。

表 1 Coll-Net 结构与输出

Tab. 1 Coll-Net structure and output

Out Size	Operator	Rescale	Add	S
$100^2 \times 32$	5×5 Conv.	No	No	2
$50^2 \times 32$	3×3 Max Pool	No	No	2
$25^2 \times 32$	3×3 Conv.	No	No	1
$25^2 \times 32$	3×3 Conv.	Yes	Yes	1
$13^2 \times 64$	3×3 Conv.	No	No	2
$13^2 \times 64$	3×3 Conv.	Yes	Yes	1
$7^2 \times 128$	3×3 Conv.	No	No	2
$7^2 \times 128$	3×3 Conv.	Yes	Yes	1
$1^2 \times 128$	GAP	—	—	—
1	FC, Sigmoid	—	—	—

3.1.2 残差网络骨架

残差块是残差网络骨架的基本组成单元而卷积层是残差块的处理单元。图 4(a)为残差块的运算过程,即输入量 x 依次通过两个卷积层变换为 $F(x)$ 后,再与 $F(x)$ 做求和计算,得到结果则输入到下一残差块进行处理。残差层卷积块的输入输出关系为:

$$\begin{aligned} F(x) &= W_2 \otimes [\sigma(W_1 \otimes x + b_1)] + b_2 \\ y &= \sigma[F(x) + x], \end{aligned} \quad (2)$$

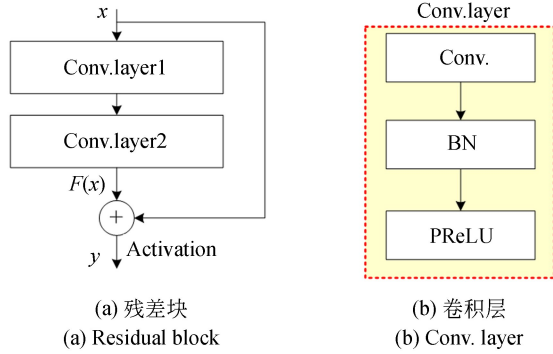


图 4 残差块和卷积层的逻辑构成

Fig. 4 Structure of residual block and conv. layer

其中: W_1, b_1, W_2, b_2 分别表示 Conv. layer 1 层和 Conv. layer 2 层的权重和偏移量; $\sigma(\cdot)$ 表示批标准化与非线性激活操作; \otimes 表示卷积操作; y 表示输出。图 4(b)为卷积层涉及的运算,即它先对输入 x 做卷积操作以提取局部特征图;然后针对同批次中各输入图对应的特征图进行批标准化处理 BN^[15],并把各神经元输入的分布拉回均值为 0 方差为 1 的标准正态分布,使后续激活函数的输入值落到函数敏感区域;最后利用激活函数对卷积层输出进行非线性化以提高网络的非线性处理能力。Coll-Net 的卷积层使用 PReLU 函数进行非线性激活,这样便能在不增加网络计算开销和过拟合风险情况下,较好地解决 ReLU 函数带来的神经元死亡问题,并有利于提升小网络的表达能力。BN 和 PReLU 的输入输出关系如式 (3),式 (4):

$$\begin{aligned} \tilde{x}_i &= \frac{\{x_i - E[x]\}}{\sqrt{\text{Var}[x]}} \\ \text{BN}_{\lambda, \beta}(x_i) &= \lambda \tilde{x}_i + \beta, \end{aligned} \quad (3)$$

$$\text{PReLU}(a_i) = \begin{cases} a_i, & a_i > 0 \\ \tau_i a_i, & a_i \leq 0 \end{cases}, \quad (4)$$

其中: x_i, \tilde{x}_i 表示批标准化输入中的第 i 个输入值与中间值; $E[x], \text{Var}[x]$ 表示输入的均值与标准差; λ, β 为可训练的批标准化加权值和偏移量; $\text{BN}_{\lambda, \beta}(x_i)$ 为批标准化结果; a_i 为激活函数输入, τ_i 为可训练的加权因子。与层数相同的普通网络相比,残差块网络可以在不增加网络参数的前提下,获得比普通网络更强的信息表征能力,其在降低障碍预测网络复杂度及提升网络预测能力方面具有很大优势。

3.1.3 空间特征重标定网络

针对行车安全系统障碍预测精度要求高的特点, Coll-Net 引入空间注意力机制对残差块进行改造,如图 5 所示。改进之处在于:采用注意力机制对第二个卷积层的结果进行通道加权,并且权重由空间特征重标定模块 (Rescale Module) 获得。加权后的空间特征按下式计算:

$$y' = \text{Padding}(\text{Scale}) \odot y, \quad (5)$$

其中: y, y' 为加权前、后的特征图, $\text{Padding}(\text{Scale})$ 表示填充后的加权值;算子 \odot 表示 Hadamard 乘积。

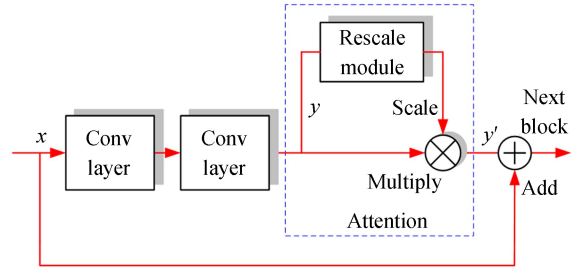


图 5 改良残差块结构

Fig. 5 Structure of improved residual block

考虑到各特征图对预测结果的贡献度不同,改良残差块采用特征重标定模块获取每个特征通道的权值 Scale,以衡量对应通道特征的重要程度,如图 6。可通过训练的方式来学习 Scale,即先令经卷积层处理后得到的通道数为 C 的特征图 y ,经过全局平均池化处理而使输出 Y 具有全局感受野;然后将其送入有 C/n 个神经元的全连接层 (FC1);再令其经过 PReLU 以提升重标定网络的非线性表达能力,进而与具有 C 个神经元的全连接层 (FC2) 相连;最后通过 Sigmoid 函数获得归一化的加权值 Scale。

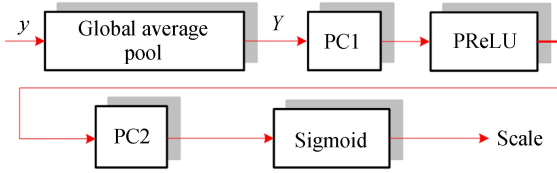


图 6 重标定模块结构

Fig. 6 Structure of rescale module

重标定模块的计算公式为:

$$Y(C_i) = \frac{1}{W \cdot H} \sum_{k=1}^W \sum_{l=1}^H y(k, l, C_i)$$

$$\text{Scale} = \text{Sig}\{\omega_2 \cdot \text{PReLU}[\omega_1 Y + \varepsilon_1] + \varepsilon_2\}, \quad (6)$$

其中 y, Y, Scale 分别表示输入的空间特征图、全局池化输出和重标定获得的加权值; W, H 表示空间特征图的宽和高; C_i 表示第 i 个通道; $\omega_1, \omega_2, \varepsilon_1, \varepsilon_2$ 分别表示 FC1 和 FC2 层的权重及偏移量; $\text{PReLU}, \text{Sig}(\cdot)$ 表示 PReLU 和 Sigmoid 函数, 并且 Sigmoid 函数的数学表达式为:

$$\text{Sig}(\alpha) = 1/[1 + \exp(-\alpha)]. \quad (7)$$

重标定模块采用两个全连接层, 旨在更好地拟合各通道间相互关系。为了降低计算复杂度, 将 FC1 层的通道数从原空间特征图的 C 降至 C/n , FC2 层的通道数继续保持为 C 。改良残差块的输出为:

$$\text{Add} = x + y'. \quad (8)$$

3.1.4 输出网络

先将第 3 个改良残差块提取的 128 张特征图进行全局平均池化处理, 然后将其送入全连接层以拟合特征图与预测值的线性关系, 最后采用 Sigmoid 生成归一化的障碍预测值。输出网络采用全局池化与全连接层组合构式, 这样既可减少特征图直接连接全连接层时产生的大量模型参数, 又可降低全连接层过拟合的危险。

3.2 训练策略

3.2.1 障碍数据集

本文结合行车实际, 依据驾驶员视觉内障碍物由远及近的变化以及应用场景中可能出现汽车、自行车、行人、电线杆、护栏、墙壁等障碍物, 构建适用于 Coll-Net 网络的数据集。该数据集基于 UZH 开放数据集 Driving Event Camera Data-

sets(DECA)^[16-17]。本文从中挑选由远及近的障碍物图片(尺寸 640×480), 并将其中可能发生车辆碰撞的图片标记为 1, 无碰撞可能的图片标记 0, 在此基础上按比例对训练集、测试集和验证集进行划分。Coll-Net 要求输入 200×200 的单通道图, 故须对数据集做预处理。为此将数据图片二值化并从中心向四周截取尺寸为 480×480 的图块, 再将其缩小为 200×200 标准图片, 作为最终实验用数据。

3.2.2 损失函数

Coll-Net 的输出经 Sigmoid 函数处理后取值处于 $[0, 1]$ 之内, 该输出视作输入图被判定为正类(有障碍)的概率, 因此训练中以二分类交叉熵函数作为基本损失函数。该函数数学形式如式(9):

$$\text{BCE}_k = -\log[p(P_k | \Pi\alpha_k)] = -P_k \log \tilde{P}_k - (1 - P_k) \log(1 - \tilde{P}_k), \quad (9)$$

其中: BCE_k 为第 k 个样本的交叉熵; P_k, \tilde{P}_k 表示第 k 个样本的真实值和预估值; $\Pi\alpha_k$ 表示第 k 个样本的输入。考虑到训练样本中负样本数远大于正样本, 故为平衡训练网络对正、负样本预测能力, 在进行整体损失函数设计时, 应使正样本的交叉熵损失权重大于负样本。本文针对正、负样本施以不同权重的损失函数定义为:

$$\text{Loss}_k = \begin{cases} -\mu \log \tilde{P}_k, & P_k = 1 \\ -(1 - \mu) \log(1 - \tilde{P}_k), & P_k = 0 \end{cases}, \quad (10)$$

其中 μ 表示正样本的权重。

3.2.3 困难样本挖掘

训练过程中大量易分类的简单样本不仅影响预测模型的优化方向, 而且影响模型对难分类样本的预测能力。本文提出降低简单样本的交叉熵损失的权重, 使模型训练更注重困难样本分类。依据真实值 P_k 与预测值 \tilde{P}_k 之差可辨识出简单样本与困难样本。引入困难样本挖掘策略的样本损失值由式(11)计算:

$$\text{Loss}_k = \begin{cases} -\mu (1 - \tilde{P}_k)^\gamma \log \tilde{P}_k, & P_k = 1 \\ -(1 - \mu) \tilde{P}_k^\gamma \log(1 - \tilde{P}_k), & P_k = 0 \end{cases}, \quad (11)$$

式中 γ 为困难样本权重调节因子,当加大 γ 时样本损失权重将降低。网络训练损失函数最终为:

$$Loss = \frac{1}{N} \sum_{i=1}^N Loss_k, \quad (12)$$

式中 N 表示样本总数。

4 基于障碍预测的行车策略

4.1 车速控制策略

基于 Coll-Net 输出的行车障碍概率,本文提出一种车速控制方法,即当 Coll-Net 对输入的车前图像做出概率预测后,处理器便开始基于预测概率 p 进行驾驶分析:当 $p=0$ 时,因无发生碰撞的可能故可保持正常行车速度;但当 $p=1$ 时,因会发生碰撞故应将车速降至 0。循此思路,本文构建出基于碰撞概率的实时控制车速模型如式 (13) 所示:

$$V_{k+1} = (1 - p_k) [(1 - \rho)V_k + \rho V_{norm}], \quad (13)$$

其中: V_k, V_{k+1} 为 $k, k+1$ 时刻行车速度; V_{norm} 为正常行车速度; $\rho \in (0, 1)$ 表示增速控制系数,当 ρ 较大时车速增长较快,反之则车速增加较慢; p_k 为 k 时刻预测碰撞概率,当碰撞概率较小时,系统以 $(1 - p_k)(V_{norm} - V_k)$ 速度提速直至正常行车速度;当碰撞概率较大时则令系统逐渐减速至 0。此法不仅为行车避障提供参考速度,而且为推进 Coll-Net 在无人车和自动驾驶上的应用提供一定的支撑。

4.2 障碍方向判定

为提升行车系统的避障能力,提出基于多窗口 Coll-Net 预测的障碍方向判定策略,以控制行车方向。该策略的思路是:先将摄像头捕获的图像尺寸缩小为 640×480 ,并在图像上方截取尺寸为 640×400 的感兴趣区域;然后在该区域的左、中、右三个区域分别截取尺寸 400×400 的窗口;再将三个窗口图像转变为尺寸 200×200 的灰度图,并将其输入到 Coll-Net 中进行障碍预测。值得一提的是,每个窗口的预测值 $pred$ 均表征对应车前方向出现障碍的概率,并且当 $pred \geq 0.5$ 时可判定该方向出现障碍物。图 7 给出了多窗口 Coll-Net 预测障碍方向的流程。

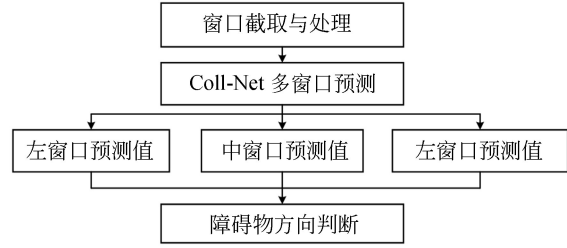


图 7 障碍物方向定位流程

Fig. 7 Process of obstacle direction localization

5 实验结果分析

5.1 实验准备

5.1.1 实验数据与平台

本文基于 DECD 构造网络训练集(含 28 806 张训练图和 1 590 张验证图)而对 Coll-Net 进行训练;使用基于 DECD 构建的测试集(含 1 576 张测试图)对 Coll-Net 的障碍预测能力进行评估;而且将实验结果与同等训练条件下采用 VGG-16^[18], ResNet, EfficientNet-B0, InceptionV3^[19] 等大型深度学习网络及谷歌专门针对嵌入式设备设计的目标检测网络 MobileNetV3^[20-21] 得到的结果做对比。特别是为了验证 Coll-Net 的障碍预测及其行车策略的有效性,文中还在 Udacity 开放的 Self-Driving 无人驾驶数据集中选取具有代表性的行车场景进行模拟实验。全部输入图片均采用与训练数据预处理相同的前处理方式。训练及测试代码均运行于 Linux 系统并使用 Intel (R) Xeon (R) W-2133 @ 3.60 GHz 六核 CPU, 至于 Coll-Net 网络的搭建则使用以 Tensorflow 作为后端的 Keras 框架且使用 opencv 进行图像预处理,且使用 python 作为编程语言。训练时使用 Nvidia RTX2080Ti 进行计算加速,考虑到嵌入式设备的计算能力,故测试中只使用 CPU 单核进行模型评估。

5.1.2 训练过程

当对 Coll-Net 模型进行训练时,使用了 Adam 优化器并将学习率设为 0.000 1,将批样本数设定为 64,正样本权重 μ 设为 0.75,训练迭代轮数 epoch 设定为 50 次。为增强预测模型的泛化能力,训练时在输出网络的全连接层设置 Dropout 为 0.4。图 8 给出了训练集损失值 $train_$

loss、准确率 train_acc 以及验证集的损失值 val_loss、准确率 val_acc 的变化。图中可见,当 $\gamma=0$ 时损失函数退化为交叉熵函数;针对训练集,当增大 γ 时训练集的模型损失值在较低轮次上得到收敛,并当 $\gamma=2$ 时预测模型不仅有较快的收敛速度而且还能保证很高的准确性;针对验证集,当 $\gamma=2$

时模型损失值和准确性具有理想的稳定性。故本文将困难样本权重调节因子 γ 设定为 2。

5.2 Coll-Net 模型测试与分析

为了全面客观地评估根据本文方法训练的 Coll-Net 模型的障碍预测性能,本文测试了不同网络模型在测试集上的准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1-score、模型参数量(Parameters)、浮点运算次数(FLOPs)和计算耗时(Latency)的性能指标,计算式分别为:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Precision = TP / (TP + FP)$$

$$Recall = TP / (TP + FN)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall}, (14)$$

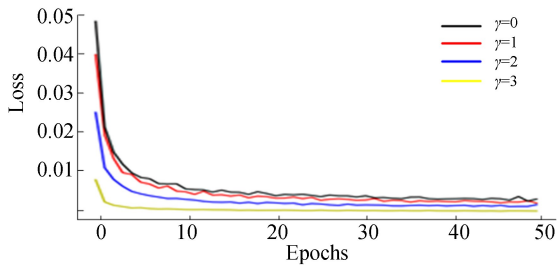
其中:TP, TN 分别表示预测正确的正样本和负样本数,FP, FN 表示预测为正而实际为负的样本及预测为负而实际为正的样本。测试中将 Coll-Net 输出概率 $prediction \geq 0.5$ 的样本记为预测正样本,其他的记为预测负样本。表 2 列出了所得到的 Coll-Net 性能指标。

表 2 模型测试指标和运算性能

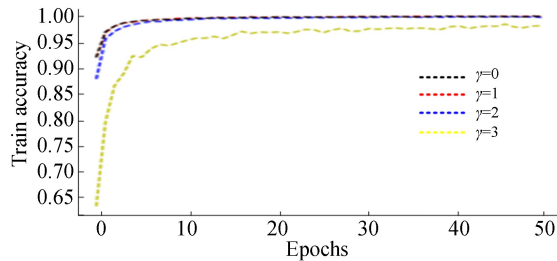
Tab. 2 Model test performance and calculated capability

Model	Accuracy /%	F1-score	Params. / ($\times 10^6$)	FLOPs / ($\times 10^9$)	Latency /ms
VGG-16	96.64	0.922	7.82	6.95	279
ResNet-18	94.35	0.875	11.53	3.01	116
ResNet-34	94.59	0.878	22.20	6.33	213
InceptionV3	92.01	0.824	23.94	4.32	167
MobileNet V3-Large	96.39	0.917	3.93	0.39	82
MobileNet V3-Small	94.44	0.870	1.52	0.09	39
EfficientNet-B0	96.26	0.914	4.05	0.61	106
Coll-Net	96.01	0.915	0.47	0.08	24

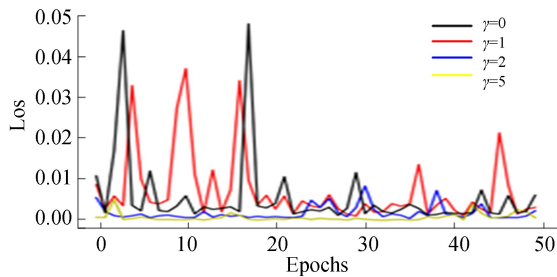
Coll-Net 在障碍预测准确率和 F1-score 两个指标上的表现稍逊于 VGG-16 (- 0.63%, -0.07) 和 MobileNetV3-Large (- 0.38%, -0.02),但它对处理器运算能力的要求(参数量和浮点运算数)远低于 VGG-16 和 MobileNetV3-Large,特别是 Coll-Net 的 24 ms 预测耗时完全达到了实时处理要求,足以适配车载嵌入式处理器



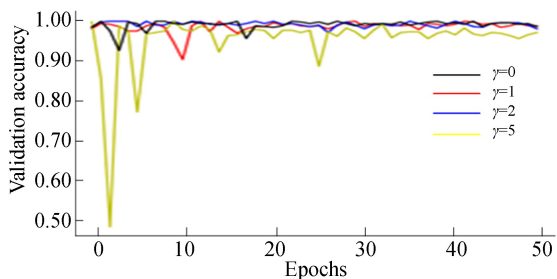
(a) 训练集损失值
(a) Loss on training set



(b) 训练集准确率
(b) Accuracy on training set



(c) 验证集损失值
(c) Loss on validation set



(d) 验证集准确率
(d) Accuracy on validation set

图 8 训练过程中的损失率与准确率

Fig. 8 Loss and accuracy on training process

计算性能。图 9 是 Coll-Net 在测试集上的预测结果混淆矩阵,图中的 Prediction 和 True 分别表示预测值与真实值,Pos, Neg 表示正、负并分别代表有碰撞可能和无碰撞可能。

图 9 可见,系统对测试集中 341 个正样本和 1 172 个负样本进行了正确预测,但将 12 张正样本误判为负(占正样本数的 3.4%)并将 51 张负样本误判为正样本(占负样本数的 4.2%)。该结果表明 Coll-Net 在障碍预测中倾向于危险判定。

此外,本文采用受试者操作特性曲线(Receiver Operating Characteristic Curve, ROC)及其对应面积(Area under the curve, AUC)作为衡量指标来对比 Coll-Net 与前文所述 Giusti, Richter 等研究者提出的障碍预测方法,如图 10 所示,Coll-Net 相比类似算法在 ROC 和 AUC 指标上取得了更好的效果。

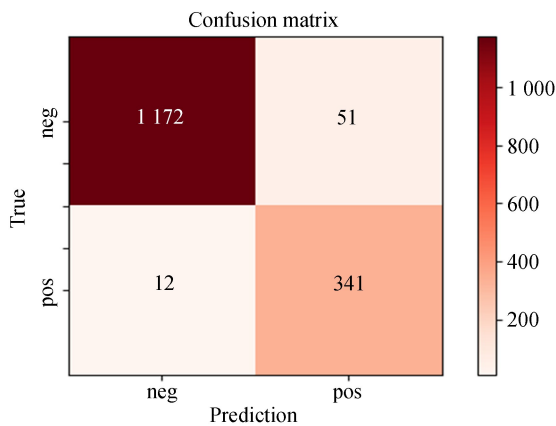


图 9 测试集上 Coll-Net 预测结果
Fig. 9 Coll-Net prediction result on test set

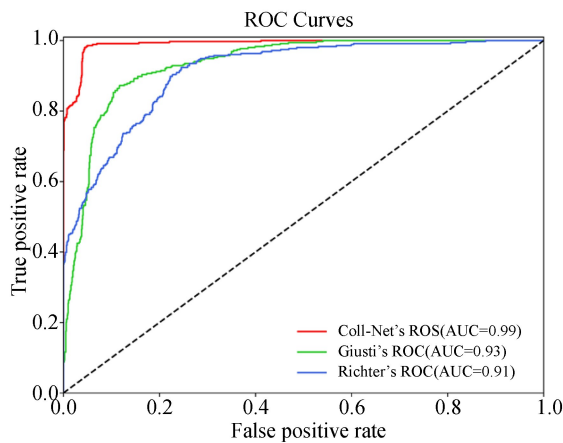


图 10 不同障碍预测方法的 ROC 曲线
Fig. 10 ROC curves of different obstacle method

为了反映 Coll-Net 的空间敏感性,本文采用可视化方法获取了在 Udacity 数据集上进行 Coll-Net 障碍预测的激活图(Classification Activation Map, CAM)^[22]如图 11、图 12(彩图见期刊电子版)。图中红色加深部分代表 Coll-Net 重点关注区域。图 11 为图像采集车在十字路口前逐渐靠近前车的 CAM,当采集车逐渐靠近前车时,网络注意力集中在前车与道路明暗差异明显的区域,这表明 Coll-Net 网络对明暗变化具有较强的敏感性。

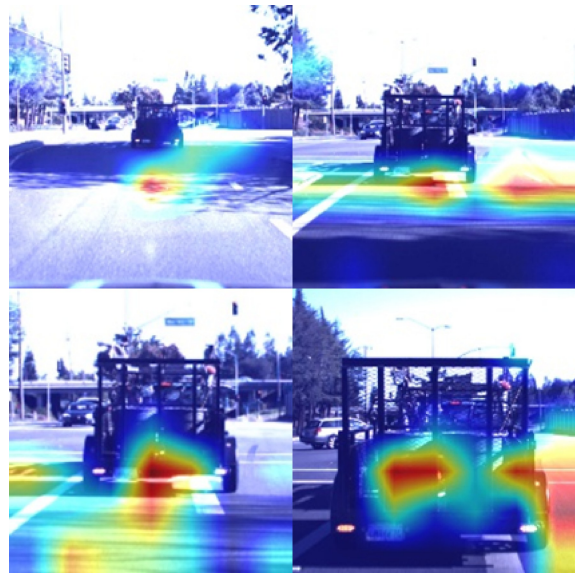


图 11 Coll-Net 十字路口场景测试
Fig. 11 Coll-Net test on crossroad

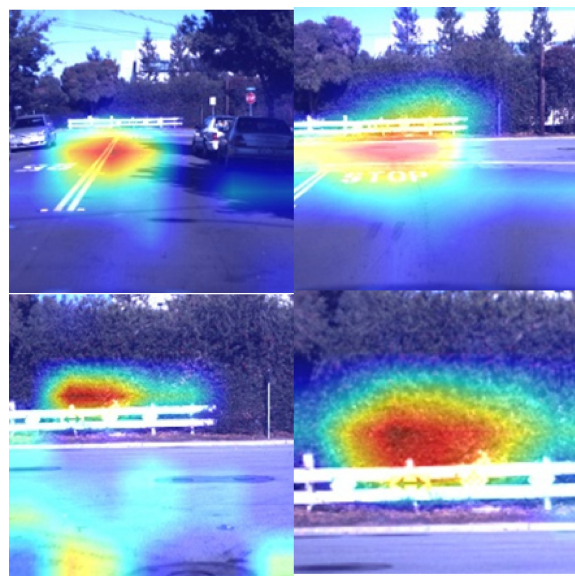


图 12 Coll-Net 丁字路场景测试
Fig. 12 Coll-Net test on T-shaped road

图 12 为图像采集车在丁字路前逐渐向树墙行进时的 CAM 图。当采集车逐渐靠近障碍物时,网络的关注区域从道路转移到障碍物本身。综合两图结果可知,Coll-Net 网络能通过道路与障碍的明暗差异以及障碍自身的特征进行障碍预测,具有广泛的行车场景适应性。

5.3 障碍预测和行车策略测试

为检验 Coll-Net 对其它障碍物及在不同光照条件下的预测能力,分别选取道路护栏、行人和低对比度场景进行测试。图 13 和图 14 分别为车辆逐渐靠近护栏和行人的预测结果,图中横坐标表示图片出现顺序,纵坐标为预测值。当障碍物较远时,预测值处于 0~0.2 的较低水平;当障碍物距离逐渐减少时预测值不断增长,最终达到 1,这表明 Coll-Net 对行人和护栏有较好的预测能力。图 15 为低对比度条件下在某十字路口前行驶图像,在采集车与前车相距较远、逐渐靠近和前车启动远离的过程中预测值则先保持不变,然后逐渐增长为 1 并保持,最后下降至 0 附近。这表明 Coll-Net 在低对比度条件下仍有良好预测能力。

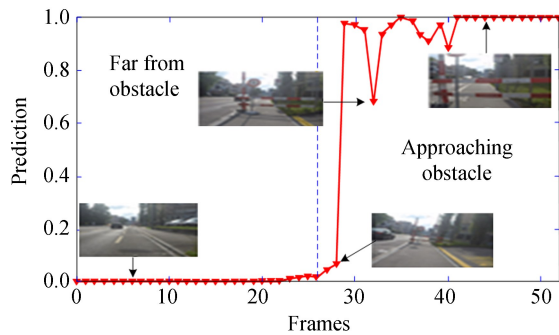


图 13 护栏障碍测试

Fig. 13 Guardrail obstacle test

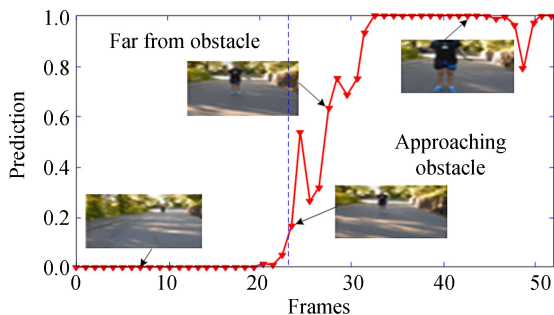


图 14 行人障碍测试

Fig. 14 Pedestrian obstacle test

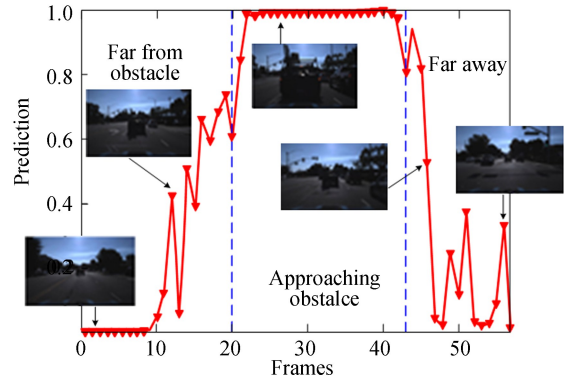


图 15 低对比度场景测试

Fig. 15 Low-contrast condition test

为了验证本文车速控制策略的有效性,在实验过程中,依据式(12)速度模式并设置增速系数 $\rho=0.5$ 且使用 Udacity 连续图像做车速计算,得到图 16 所示结果。图中,Pred 为 Coll-Net 预测值,Speed 表示控制车速与正常车速 V_{norm} 的比值。可见,当车距减少时则车速随预测值增大而迅速减小,这表明了本文基于 Coll-Net 的车速控制策略的有效性。

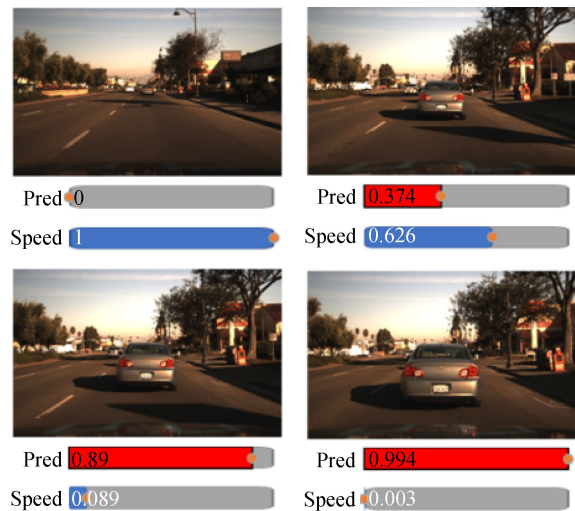


图 16 车速控制策略测试结果

Fig. 16 Speed control policy test

图 17 是利用多窗口 Coll-Net 预测障碍方向的结果(彩图见期刊电子版)。当将左、中、右三个窗口的图像输入 Coll-Net 后得到对应的窗口预测值,将预测值 $p \geq 0.5$ 的窗口用红色定位框

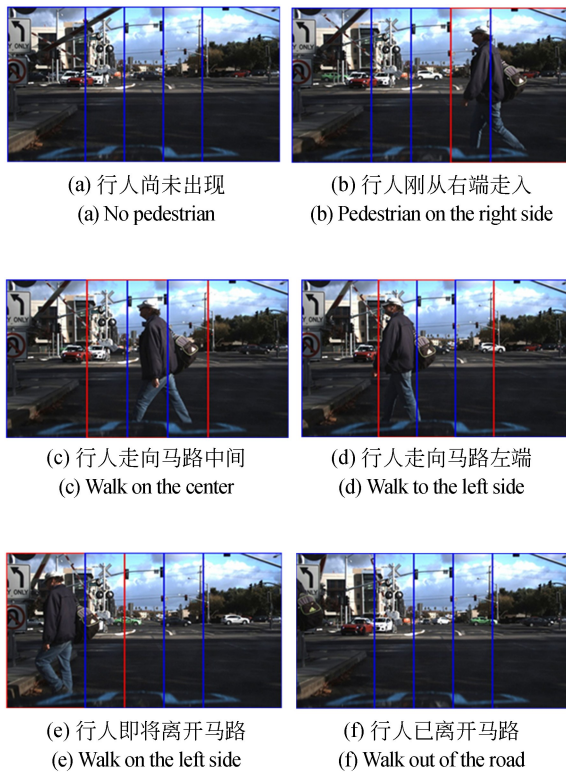


图 17 障碍方向定位测试结果

Fig. 17 Result of obstacle direction localization

参考文献:

- [1] A HASELHOFF, A KUMMERT. A vehicle detection system based on Haar and triangle features [C]. 2009 *IEEE Intelligent Vehicles Symposium*, 2009: 261-266.
- [2] 刘峰,王思博,王向军,等. 多特征级联的低能见度环境红外行人检测方法 [J]. *红外与激光工程*, 2018, 47(6): 127-134.
LIU F, WANG S B, WANG X J, *et al.*. Infrared pedestrian detection method in low visibility environment based on multi feature association [J]. *Infrared and Laser Engineering*, 2018, 47(6): 127-134. (in Chinese)
- [3] J REDMON, A FARHADI. YOLO9000: Better, faster, stronger [C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017: 6517-6525.
- [4] S REN, K HE, R GIRSHICK. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [5] 范丽丽,赵宏伟,赵浩宇,等. 基于深度卷积神经网络的目标检测研究综述[J]. *光学精密工程*, 2020, 28(5): 1152-1164.
FAN L L, ZHAO H W, ZHAO H Y, *et al.*. Survey of target detection based on deep convolutional neural networks [J]. *Opt. Precision Eng.*, 2020, 28(5): 1152-1164. (in Chinese)
- [6] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 39(4): 640-651.
- [7] 任凤雷,何昕,魏仲慧,等. 基于DeepLabV3+与超像素优化的语义分割[J]. *光学精密工程*, 2019, 27(12): 2722-2729.
REN F L, HE X, WEI ZH H, *et al.*. Semantic segmentation based on DeepLabV3+ and superpixel optimization [J]. *Opt. Precision Eng.*, 2019, 27(12): 2722-2729. (in Chinese)
- [8] 王中字,倪显扬,尚振东. 利用卷积神经网络的自动驾驶场景语义分割[J]. *光学精密工程*, 2019, 27(11): 2429-2438.
WANG ZH Y, NI X Y, SHANG ZH D. Autonomous driving semantic segmentation with convolu-

在原图中进行标记,以表明该方向出现障碍。当行人分别在车辆的右方、中央和左方出现时, Coll-Net 都能准确判断了行人所在的区域,这也表明了本文障碍方向预测策略具有可行性。

6 结 论

本文提出了融合空间注意力机制的障碍预测网络 Coll-Net 用于各类行车障碍的预测。Coll-Net 网络以单目图作为输入,利用空间注意力机制改良的残差块网络提取图像的重要特征,通过全连接层输出预测的障碍概率,解决了基于目标检测的障碍预测算法中的目标单一性问题,对各类典型障碍物均具有预测能力;与现有算法相比较大提升了预测的准确性,并且能根据多窗口方式大体定位障碍物方向。Coll-Net 在 DECD 数据集上准确率达到 96.01%, F1-score 达到 0.915,模型推理时间仅需 24 ms,达到处理的准确性和实时性需求,为 Coll-Net 在智能驾驶的应用上奠定了一定的技术基础。

- tion neural networks [J]. *Opt. Precision Eng.*, 2019, 27(11): 2429-2438. (in Chinese)
- [9] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[J]. *Advances in neural information processing systems*, 2012, 25(2): 1097-1105.
- [10] TAN M, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[C]. *36th International Conference on Machine Learning (ICML), California*, 2019: 691-700.
- [11] GIUSTI A, GUZZI J, CIRESAN D, *et al.*. A machine learning approach to visual perception of forest trails for mobile robots[J]. *IEEE Robotics & Automation Letters*, 2016, 1(2):661-667.
- [12] C RICHTER, N ROY. Safe visual navigation via deep learning and novelty detection[C]. *Robotics: Science and Systems XIII (RSS), Massachusetts*, 2017:64-73.
- [13] 刘玉洁, 朱韶平. 基于全局和局部多特征的图像增强算法[J]. *液晶与显示*, 2020, 35(5): 508-512.
- LIU Y J, ZHU SH P. Image enhancement algorithm based on global and local multi features[J]. *Chinese Journal of Liquid Crystals and Displays*, 2020, 35(5): 508-512.
- [14] K HE, X ZHANG, S REN, *et al.*. Deep residual learning for image recognition[C]. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV*, 2016: 770-778.
- [15] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]. *Proceedings of the 32nd International Conference on Machine Learning, Lille*, 2015: 448-456.
- [16] LOQUERCIO A, SEG? M, SCARAMUZZA D. A general framework for uncertainty estimation in deep learning[J]. *IEEE Robotics and Automation Letters*, 2019, 5:3153-3160.
- [17] LOQUERCIO A, MAQUEDA A I, DEL-BLANCO C R, *et al.*. DroNet: learning to fly by driving [J]. *IEEE Robotics & Automation Letters*, 2018, 3(2):1088-1095.
- [18] K SIMONYAN, A ZISSERMAN. Very deep convolutional networks for large-scale image recognition[C]. *International Conference on Learning Representations (ICLR), San Diego*, 2015: 1-14.
- [19] C SZEGEDY, V VANHOUCHE, S IOFFE, *et al.*. Rethinking the inception architecture for computer vision[C]. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV*, 2016: 2818-2826.
- [20] M SANDLER, A HOWARD, M ZHU, *et al.*. MobileNetV2: Inverted residuals and linear bottlenecks [C]. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT*, 2018: 4510-4520.
- [21] A HOWARD, M SANDLER, G CHU, *et al.*. Searching for mobileNetV3[C]. *2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South)*, 2019: 1314-1324.
- [22] B ZHOU, A KHOSLA, A LAPEDRIZA, *et al.*. Learning deep features for discriminative localization[C]. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV*, 2016: 2921-2929.

作者简介:



雷俊锋(1975—),男,湖北武汉人,博士,副教授,2002年于武汉大学电子信息学院获得博士学位,主要研究方向是图像处理与人工智能。E-mail:jflel@whu.edu.cn



贺睿(1997—),男,江西南昌人,硕士研究生,主要研究方向是图像处理与智能感知,E-mail:he_rui@whu.edu.cn